Data Solutions and a Platform for use to Meet UN Sustainable Development Goals by 2030

Steven Hagan
Vice President Engineering
Oracle Database Server Technologies
August, 2017
PLATFORM: COMPUTING / STORAGE TRENDS:

- Computer System Performance –
  - Hardware - **EVOLUTIONARY** – Moore’s law still holding
  - New possibilities at Research Level – not yet proven
    - DNA for Storage; 3D Glass, Holography; Carbon Nanotubes, Graphene, Quantum
  - Software – **DISRUPTIVE** – Parallelism => clusters of **10,000+** computers, CLOUD, ML, AI

- Software – **FLEXIBILITY** - **NOW** Supporting many Data types in Databases
  - Databases/persistent stores: **POLYGLOT PERSISTENCE** now can handle **ALL** types of data
  - Software – **GRAPH STORAGE, SEMANTICS, ONTOLOGIES**
    - – Add all types of data, build **NEW** relationships
  - Enables **MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE** (ML, AI)
  - Stream data arriving; Filter the data; ML: Keep what matches your requirements; aggregate it, make it accessible for **ALL SEVENTEEN (17)** goals.
Acquiring/Keeping Data for **17** Sustainable Development Goals: Need One Platform for **ALL** Variety, Velocity, Volume of Data

- VIDEO: UAVs, DRONES, SURVEILLANCE
- IMAGERY/Raster: (Satellites, Medical)
- Sensors (IOT), LIDAR, 3D, RFID, Wearables
- Social Media, Web Scraping, Mobile Phones
- New data products for: Land and Water mgmt, Agriculture, Environment Transportation, Terrain and City Models, SDIs for planning, maintenance, Emergency response, Defense, Intelligence, Consumers, Healthcare
- Genomics (DNA Sequencing)
- Semantics, Ontologies
- Machine Learning, AI
- Location is a Powerful Organizing Principle
- MULTIPLE VERSIONS OF THE ABOVE
Managing All Spatial, Graph, Statistic Data – in One Store

Location and Statistics analysis with Secure, scalable storage for enterprise data

Oracle Spatial and Graph

SQL access
Web Services (OGC)
SPARQL End Point

40+ Graph Analysis Functions (PGX)

Property Graphs
Network Graphs
RDF Semantic Graphs
Topologies

Spatial Vector Acceleration
Geocoding
Routing
Inferencing

Rasters
3D, point clouds (LiDAR)
Oracle Statistics / Analytics Machine Learning Algorithms

- **Classification**
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - Neural Network
  - Support Vector Machine
  - Naïve Bayes
  - Explicit Semantic Analysis
  - Gaussian Mixture Models

- **Clustering**
  - Hierarchical K-Means
  - Hierarchical O-Cluster
  - Expectation Maximization

- **Anomaly Detection**
  - One-Class Support Vector Machine

- **Regression**
  - Generalized Linear Model
  - Support Vector Machine
  - Random Forest
  - Linear Model
  - Stepwise Linear regression
  - LASSO

- **Association Rules**
  - A priori

- **Attribute Importance**
  - Minimum Description Length
  - Principal Component Analysis
  - Unsupervised Pairwise KL Divergence

- **SQL Predictive Queries**

- **Statistical Functions**

- **Algorithm Text Support**
  - Algorithms support text type
  - Tokenization and theme extraction
  - Document similarity

- **Feature Extraction**
  - Principal Component Analysis
  - Non-negative Matrix Factorization
  - Singular Value Decomposition

- **Time Series**
  - Single Exponential Smoothing
  - Double Exponential Smoothing

- **Open Source ML Algorithms**
  - CRAN R Algorithm Packages through Embedded R Execution
  - Spark MLLib algorithm integration
Spatial: Open and interoperable
Sustainable Goals: Repurposing Data: Ontology-driven
Enable Shared, Actionable Knowledge

- Simple Features
- GeoRaster
- Topology
- Networks
- Gazetteers

- Data Integration
- National Map schemas
- Geographic names
- Temporal
- Naïve Geography
- ...

Application Ontologies

Environmental Monitoring
Disaster Recovery
Healthcare Biotech

National Mapping Private Cloud

Spatial Data

Geographic Names

Raster Data

RDF & OWL Metadata

Environmental Monitoring
Disaster Recovery
Healthcare Biotech
Harmonizing the Electronic Health Care Ecosystem – Goal 3 Using Semantics, Ontologies

Enterprise-wide, Patient-centric, longitudinal Record System

Domain Ontologies
(business metadata + Ontologies)

Data Servers

Lab/clinical Care
Research
Content Mgmt
Billings/Claims
Reporting/BI

Data Sources / Data Types

Social Media
Medical Devices
Lab Information Systems
Subscription Services
Legacy Patient Records
Linked Open Data: Connecting With other Services and Clouds
Oracle: Linked Data support: on-premise or in the Cloud

• Highly scalable, secure triple store based on RDF (Resource Description Framework)
  – **1 TRILLION TRIPLE BENCHMARK**, leading Triple Store:W3.org
    – 1.13 million triples per second query performance

• SPARQL and SPARQL in SQL support
  – Apache Jena and OpenRDF Sesame pre-integrated
  – SPARQL endpoint enhanced with query control
  – GeoSPARQL support (classes, properties, datatypes, query functions)

• Forward-chaining based inferencing engine in the database
  – Various native rulebases (RDFS, OWL2 RL, SKOS, ...), integration with OWL2 reasoners (TrOWL, Pellet)

• RDB to RDF mapping on relational data aligned with RDB2RDF standard
Support Breadth of National & UN Data

ABOVE STOYPEPES

Data arrives, is filtered, stored data is available to ALL Organizations

GUIDANCE: THIS IS AN ARCHITECTURE TO SUPPORT ONE SHARED MULTIPURPOSE NATIONAL / UN STORE
You Enhance Innovation & Sharing By Using **STANDARDS**
e.g. – The Spatial / Semantics Data Domains

- **ISO**
  - TC 211; TC 204

- **Open Geospatial Consortium**
  - Simple Features; GML; Web Services

- **De-facto Standards**
  - SHP, MGE, DXF, KML

- **Professional Standards**
  - ISPRS, FIG, WMO

- **Java, .NET, Flash**

- **W3C: RDF, OWL, SPARQL, GeoSPARQL**

- **TAGGED METADATA** – agree on tags

---

**SQL3/MM Spatial**
Public Clouds, Private Clouds: Data/Statistics Platforms

- Used by multiple tenants on a shared basis
- Hosted and managed by cloud service provider

Public Clouds

- SaaS
- PaaS
- IaaS

Private Cloud

- Apps
- PaaS
- IaaS

Trade-offs

Lower *upfront* costs ↔ Lower *total* costs
Outsourced management ↔ Greater control over security, compliance, QoS
OpEx ↔ CapEx & OpEx
ELASTICITY is key value of Clouds

Oracle Technology Supplies both Public and Private clouds

---
YOU MAY NEED A CLOUD IN EACH COUNTRY ---DEPENDS ON THEIR LAWS
To Meet 2030 Goals: Do Not Build Your Solutions From Scratch
Long Term Cost of Ownership rises with custom construction & Open Source

UN-GGIM: “train the individuals is at least five years”
Sustainable Goals: All Data Types /Ontologies/ ML / AI Bases: Success Enhanced with **MULTI-MODEL DATABASE PLATFORM**

**Big & Fast Data**
- Volunteered Geographic Statistical Information
- Sensors
- Streaming Data
- Geo-referenced Video, 3D, LiDAR Satellites

**Simplify Statistics IT**
- Support for Open Standards
- Spatial Database, Application Server, BI, tools
- Support by Leading Partner solutions
- Multi-Model Engineered Systems

**Deep Analytics**
- Real-time Complex Event Processing
- Dense Visualization
- Spatial Analysis Graph Analytics

**On Premise, On Cloud, Shared Services**
- Shared GeoSpatial Services
- Location Aware Everything

**Fully Parallel and Secure**